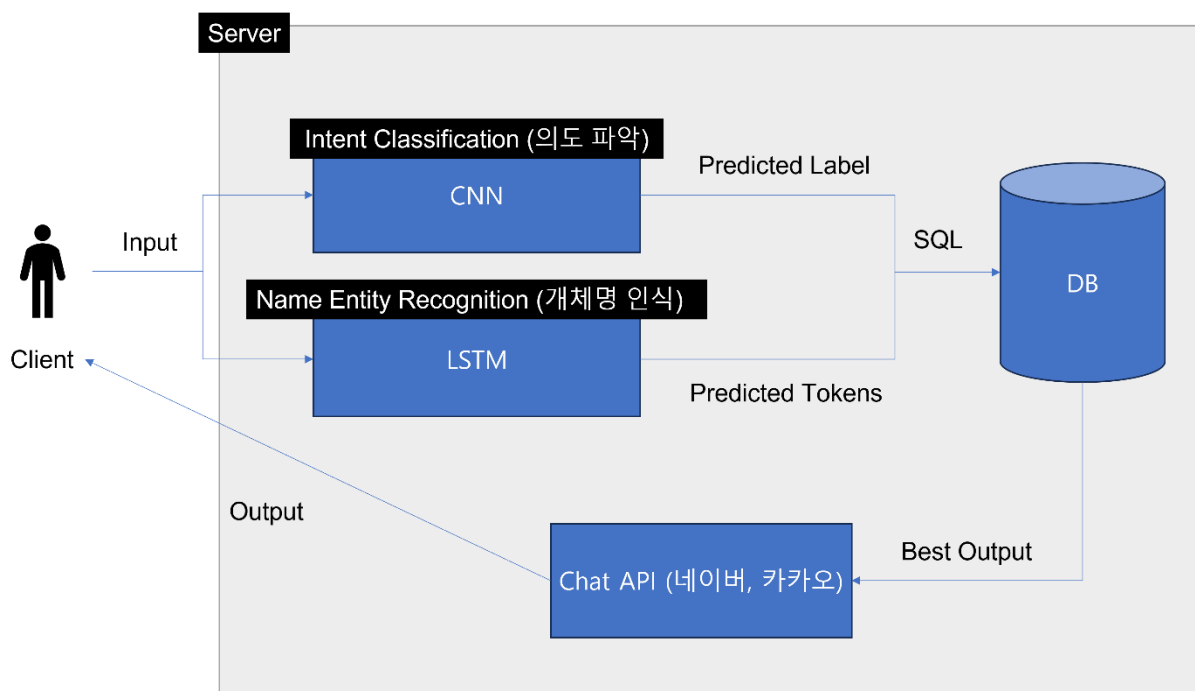


[수행 내용]

1. 프로젝트 구조 이해 및 정리

프로젝트 구조를 시각화 하면 다음과 같습니다.



프로젝트 진행상황은 철도연구원 FAQ 데이터로 CNN, LSTM 모델을 학습시키는 것까지 진행되었습니다. 현재 개발이 필요한 부분은 다음과 같습니다.

- 1) 답변 도출 모델 (CNN, LSTM 모델 결과를 기반으로)
- 2) DB 설계 (Client의 질문 데이터를 저장하는 DB, 답변 데이터가 저장되어 있는 DB) (현재는 엑셀 사용)
- 3) Chat API를 이용한 챗봇 엔진 및 서버 개발

2. Local Project를 Colab 환경으로 이전

Local Project를 Colab 환경으로 이전하여 접근성을 향상시켰습니다. Colab이 갖는 이점은

다음과 같습니다.

- 1) 공간 제약 없이 코딩 및 결과 확인 가능
- 2) 프로젝트 참여자들이 모두 동일한 환경으로 쉽게 디벨롭 및 코드 실행 가능

▼ QA Chatbot

Structure

{개발중: 완료 X, 디벨롭 필요: 완성이[보수 필요], 주석없음: 완성 O}

chatbot

```
├ config
├ train_tools
│   └ dict
├ utils
├ model
│   ├── intent
│   └ ner
├ test
└ sample
```

이메일에 해당 colab 파일 함께 첨부하여 보내드립니다.

3. 개발 목표 및 애로사항 파악

현재 개발 목표는 철도연구원에서 사용가능한 Q&A 챗봇을 만드는 것입니다. 프로토타입은 8월 중순 이전까지 개발 완료를 목표로 하고 있습니다. 프로토타입 답변의 일련의 과정은 다음과 같습니다.

채팅 플랫폼에서 제공하는 API를 바탕으로 클라이언트와 상호작용이 가능한 서비스를 배포합니다. 이후 클라이언트의 질문을 받아 CNN모델로 Indent(문장의 의도)를 파악하고, LSTM 모델로 NER (개체명 인식)을 진행합니다. 그리고 두 모델의 결과를 바탕으로 SQL 조건문을 통해 가장 적합한 답변을 답변이 저장되어 있는 DB에서 추출합니다. 해당 답변을 채팅 플랫폼으로 클라이언트에게 전달합니다.

현재 개발 애로사항은 다음과 같습니다.

- 1) 데이터 부족

금일 전체적인 프로젝트 진행상황을 파악한 결과, 학습에 필요한 데이터가 크게 부족하다는 사실을 알게 되었습니다. CNN 의도 파악 모델의 경우에는 적은 데이터에도 비교적 준수한 답변을 도출했지만, LSTM 개체명 인식 모델의 경우에는 학습이 거의 되지 않는 모습을 보였습니다.

다만, 전반적으로 모델이 개발 완료된 상황으로 추후 데이터를 보강한다면 모델의 Accuracy가 증가할 것으로 보입니다.

해당 상황에 대해서 원래 주어진 Sample 데이터를 바탕으로 이번 주 내로 답변 도출 모델까지 개발하는 것을 목표로 하고 있습니다.